Claims

1. A method of reading striped digital data from a RAID array of disk drives, each drive having a respective data port of predetermined width coupled to an internal buffer, the method comprising the steps of:

providing a single buffer memory having a data port coupled to all of the disk drive data ports for transferring digital data;

providing a single address counter for addressing consecutive locations in the buffer memory;

sending read commands to all of the disk drives so as to initiate read operations in all of the disk drives;

waiting until read data elements are ready at all of the disk drive data ports;

after read data elements are ready at all of the disk drive data ports, synchronously retrieving and storing the read data elements from all of the disk drive data ports into consecutive locations in the buffer memory under addressing control of the single address counter;

wherein said synchronously retrieving and storing the read data elements from all of the disk drive data ports includes clocking the read data through a common pipeline so as to form a contiguous word serial data stream through the pipeline;

concurrently computing redundant data from the read data while the read data moves through the pipeline;

and, if a failed drive has been identified, substituting the computed redundant data into the word serial data stream in lieu of the failed disk drive data; and

storing the word serial data stream into the buffer memory thereby providing the requested read data without incurring delay to reconstruct data stored on the failed ~~disk drive~~.

2. A RAID disk array controller comprising:

host bus interface means for interfacing to a host bus for data transfer;

buffer memory means for storing data;

a processor for controlling operation of the disk controller so as to effect synchronous data transfers between the buffer memory and an array of disk drives;

disk drive interface means including a drive data bus for interfacing the controller to an array of disk drives including a redundant drive;

redundant data operating means disposed along the drive data bus for forming redundant drive data on the fly as data passes from the buffer memory to the drives during a disk write operation.

3. A disk array controller according to claim 2 wherein the redundant data operating means includes:

a multiplexer having a first input coupled to the buffer memory port to receive write data;

an XOR/LOAD circuit having a first input coupled to the buffer memory port;

an accumulator coupled to the output of the XOR/LOAD circuit;

a feedback path from the accumulator circuit to a second input of the XOR/LOAD circuit;

the multiplexer having a second input coupled to the accumulator; and

the multiplexer output coupled to the drive data bus for interfacing to the array of disk drives, so that in operation the multiplexer selects either a word of write data from the buffer memory for writing to disk, or a redundant word formed in the accumulator for writing to disk as redundant data.

4. A disk array controller according to claim 2 further comprising second redundant data operating means disposed along the drive data bus for reconstructing missing data on the fly as data passes from the drives to the buffer memory during a disk read operation, so that in operation a single-drive failure does not cause loss of data or delay in providing requested read data to the buffer memory.

5. A disk array controller according to claim 4 wherein the second redundant data operating means includes:

a pipeline of registers through which read data is passed during a disk read operation;

an input end of the pipeline coupled to the disk drive data bus to receive read data;

a multiplexer having a first input coupled to an output end of the pipeline to receive read data;

an XOR circuit coupled to the disk drive data bus to receive read data;

an accumulator having an input coupled to the XOR circuit output;

a holding circuit having an input coupled to the XOR circuit output;

a holding circuit having an input coupled to the accumulator to hold accumulated data;

a feedback path from the output of the accumulator to a second input of the XOR circuit for forming XOR data in the accumulator as valid read data passes through the XOR circuit from the drive data bus;

an output path from the hold circuit to a second input of the multiplexer to provide reconstructed missing data;

wherein the multiplexer output is coupled to the buffer memory so that in operation, for each read strobe, the multiplexer selects either a word of valid read data from the pipeline for writing to the buffer memory, or a reconstructed word formed in the accumulator for writing to the buffer memory in lieu of missing or bad data.

6. A disk array controller apparatus comprising:

a buffer memory (106);

disk drive interface means (204) for connection to a plurality of disk drives;

a data bus (310) interconnecting the buffer memory and the disk drive interface means;

control means (1200) coupled to the buffer memory and coupled to the data bus for synchronously transferring data over the data bus between the buffer memory and the interface means to effect disk read and disk write operations, wherein the control means includes only a single DMA channel for addressing the buffer memory; and

means disposed between the buffer memory and the data bus for generating redundant check data on the fly during execution of a disk write operation.

7. A disk array controller apparatus comprising:

a buffer memory (106);

disk drive interface means (204) for connection to a plurality of disk drives;

a data bus (310) interconnecting the buffer memory and the disk drive interface means;

control means (1200) coupled to the buffer memory and coupled to the data bus for synchronously transferring data over the data bus between the buffer memory and the interface means to effect disk read and disk write operations, wherein the control means includes only a single DMA channel for addressing the buffer memory; and

means disposed between the buffer memory and the drive data bus for reconstructing missing data during a read operation so that only correct read data is stored in the buffer memory.

8. A disk array controller apparatus according to claim 7 wherein the data reconstructing means includes a pipeline of registers arranged for transferring word serial read data from the data bus to the buffer memory.

9. A disk array controller apparatus according to claim 8 wherein the pipeline includes a number of stages equal to N+1, where N is the total number of said disk drives in the array, each stage in the pipeline having a number of bits equal to a number of bits in the drive data bus.

10. A method of writing digital source data stored in a buffer to a RAID array of N disk drives numbered 0 to (N-1), each disk drive having a like drive port including a data bus of predetermined width, the method comprising the steps of:

sequentially reading the source data from a contiguous block of memory locations in the buffer, thereby forming a serial stream of source data;

selecting a data element size equal to an integer multiple of the data bus width of the drive port;

striping the source data read from the buffer by the selected data element size across the drives by writing an xth data element of the source data to drive (x mod N).

11. A method of writing to a RAID array according to claim 10 and further comprising:

providing an additional drive number N+1;

computing redundant data in response to the serial stream of source data;

writing the redundant data to the N+1 drive.

12. A method of writing to a RAID array according to claim 11 wherein said computing step includes determining a redundant data element in response to each N data elements of the serial stream of source data.

13. A method of writing to a RAID array according to claim 12 wherein the redundant data element consists of the boolean XOR function of the corresponding N data elements.

14. A method of writing to a RAID array according to claim 10 wherein the data bus width of the drive port is 16 bits.

15. A method according to claim 10 wherein the selected data element size is 16 bits.

16. A method of writing digital source data stored in a buffer to a RAID array of N+1 disk drives numbered 0 to N, comprising the steps of:

sequentially reading the source data from a contiguous block of memory locations in the buffer, thereby forming a single, serial stream of source data having a transfer rate;

synchronously forming redundant data responsive to the serial stream of data at the same transfer rate as the serial stream of data;

inserting the redundant data into the serial stream of data; and writing the resulting serial stream of data in striping fashion to the N+1 disk drives.

17. A method according to claim 16, each disk drive having a like drive port including a data bus of predetermined width, further comprising selecting a data element size equal to an integer multiple of the data bus width of the drive port; and wherein:

said step of synchronously forming redundant data includes determining a single redundant data element in response to each N data elements of the serial stream of source data;

said step of inserting the redundant data into the serial stream of data consists of inserting each redundant data element into the serial stream as a next data element immediately following the N data elements used to form the said redundant data element; and

said writing step includes striping the resulting serial stream of data, including the redundant data, by the selected data element size across the drives whereby the redundant data elements are stored on drive N+1.

18. A method according to claim 16 wherein the selected data element size is 16 bits.

19. A method according to claim 16 wherein the selected data element size is 32 bits.